Bias estimation of p-values in analytic and simulated Cox Tests for non-nested models

Paul Wilson

[1] Department of Mathematics, National University of Ireland, Galway
Paul.Wilson@nuigalway.ie

Abstract: In this paper we show that the estimation of p-values in both Cox's test for non-nested models and its simulation based analogues is biased and that whilst simulation based Cox tests may be extended to nested models the consequent level of bias is so large as to render the test useless, but that this bias may be removed by adapting the null hypothesis to be simple.

Keywords: Bias, Cox's Test, Hybrid Test

1 Introduction

One of the few tests for comparing non-nested models is the analytic test of Cox (1962), simulation based variants of which have been proposed by Williams (1970) and Hinde (1993). Cox's analytic test determines p-values for $H_0 : M_f$ *is the correct model* and $H_0 : M_g$ *is the correct model* from which we may classify the data. Whilst Hinde focusses on model discrimination, he shows that p-values may be derived from his test by standard procedure. In developing his test Cox makes various approximations, and notes that these introduce bias, which could be reduced if these approximations were not made. Whilst simulation based versions of Cox's test avoid his calculations and thus his approximations, problems arise with the estimation of p-values as the underlying test statistic, T_f, the difference between the observed log-likelihood ratio and its expected value, depends upon the parameters of the data in question. Thus the null hypotheses are composite, and biased estimation of p-values may occur.

2 Bias and the Distribution of p values

A p-value may be defined by (see Davison and Hinkley (1999)):

$$p_{\text{obs}} = Pr_0\left(T \geq t_{\text{obs}}\right) \tag{1}$$

where $T = t(Y)$ is some function of data, Y, and Pr_0 indicates probability under the null hypothesis.

Assuming that F_0, the null distribution function of T is known and continuous, we may regard p_{obs} as an instance of a random variable $\mathcal{P} = 1 - F_0(T)$, thus, for $0 \leq \xi \leq 1$:

$$Pr_0 \left[1 - F_0(T) \leq \xi\right] = Pr_0 \left[T \geq F_0^{-1}(1-\xi)\right] = 1 - F_0 \left[F_0^{-1}(1-\xi)\right] = \xi \quad (2)$$

i.e. the null distribution of $\mathcal{P}$ is the uniform distribution. Whilst the above only exactly holds if F_0 is continuous, it still holds to a great extent for discrete F_0. If F_0 is not known, as frequently occurs when the underlying null hypothesis is composite, then the distribution of the estimated p values will fail to be uniform, and the interpretation (1) will fail to hold, i.e. $E\left(\hat{p}\right) \neq p'$, where p' is the true value of p, i.e. the estimation process is biased. Reversing the argument of (2) we see that the converse also holds, i.e. if the distribution of the p-values is uniform, then their estimation is unbiased.

3 Bias in Cox's Analytic Test

Cox shows that $L_f(\alpha) - E\left\{L_f(\hat{\alpha})\right\}$ is normally distributed with mean approximately zero and a stated variance term, but proceeds to acknowledge that a correction for bias could be obtained by taking the mean to be

$$-\frac{E_\alpha\left\{f(f_\alpha^2 + f_{\alpha\alpha})\right\}}{2E_\alpha(f_\alpha^2)} - \frac{1}{2}, \quad (3)$$

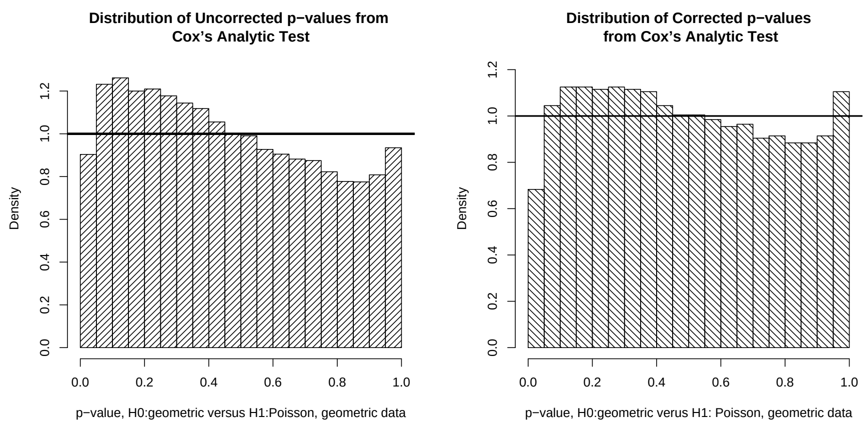
Cox (1962) makes no attempt to derive an analogue of (3) for T_f. Such an analogue requires the evaluation of complicated integrals which due to the incorporation of terms involving the expectation over M_f of functions of the "competing model" $g(y, \beta_\alpha)$, does not "reduce" to a term of the relative simplicity of (3). We may however determine the departure from zero of the mean of T_f by simulation. The second column of Table 1 summarises $E_\alpha(T_f)$ under H_0:Poisson *versus* H_1: geometric, where the data is in fact Poisson; in the third column the data are geometric and the hypotheses are reversed. These expected values are each based upon the results of $100,000$ simulations. Clearly the approximation to zero is non-negligible.

Figure 1 illustrates the distributions of unadjusted p-values and p-values that have been adjusted by taking the mean to be as in Table 1. These mean values were obtained from repeated application of Cox's test to data drawn from *geometric*(1) data, with the hypotheses H_0:*geometric* versus H_1 :*Poisson*. We see that the correction reduces, but does not eliminate, bias. Indeed the bias actually increases at the lower tail.

TABLE 1. Expected values of T_f and T_g, $n = 100$, based upon 100,000 simulations.

Mean	H_0:Poisson (Poisson Data)	H_0:geometric (geometric data)
0.6	0.15	-0.37
1.0	0.25	-0.63
1.4	0.29	-0.82
1.8	0.32	-1.14
2.2	0.36	-1.24

FIGURE 1. Distribution of uncorrected and corrected p-values geometric(1) data, H_0:geometric, H1:Poisson, $n = 100$



Bias in Simulation-Based Cox Tests

The left hand diagram of Figure 2 illustrates that bias occurs when a simulation based Cox test is used to determine p-values. To emphasise that this bias is due to the composite nature of the null hypothesis, the right hand diagram shows that if the parameters of the null hypothesis are fixed, the resultant p-values are uniformly distributed. Whilst the level of bias may be reduced by multiple bootstrapping, this is usually not practical.

It is possible to extend simulation based Cox tests to nested models, however bias in the estimation of p-values is enormous, as illustrated by Figure 3. The left-most diagram shows the distribution of the unadjusted p-values, the center diagram that of p-values adjusted by a double bootstrap, and the right-hand diagram the distribution of the p-values under the *simple* hypothesis where the model parameters are fixed.

FIGURE 2. Distribution of p-values Poisson(0.8) data, H_0:Poisson, H1:geometric, $n = 50$. Parameters varying and fixed

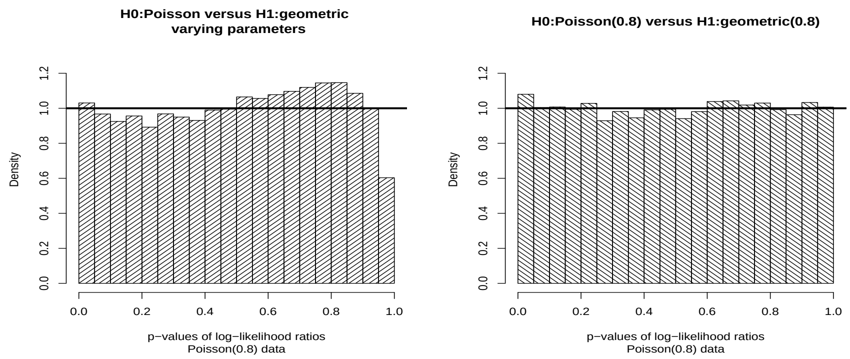
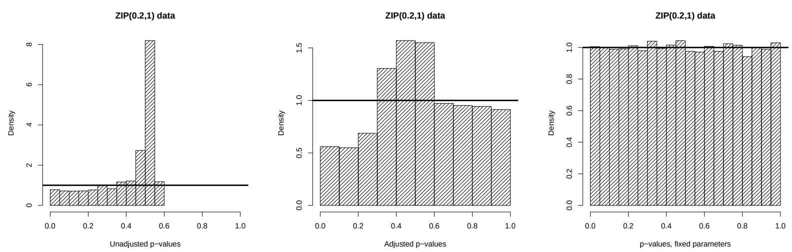


FIGURE 3. Distribution of p-values ZIP(0.2,1) versus Poisson data, H_0:Poisson, H1:geometric, $n = 50$. Parameters varying and fixed



4 Conclusion

Bias is evident in all forms of Cox's test. Bias is absent in tests based upon simple null hypotheses incorporating fixed model parameters such as the Hybrid test of Wilson (2007) and the proposed "Dragnet Test", discussed elsewhere at this conference.

References

Cox DR. (1962). Further Results on Tests of Separate Families of Hypotheses. *Journal of the Royal Statistical Society. Series B* **24**, 406–423.

Davison AC and Hinkley DV (1999), Bootstrap Methods and their Applications, 3rd ed. *Cambridge University Press.*

Hinde JP. (1992). Choosing Between Non-nested Models: a Simulation Approach. In Fahrmeir L et al. eds. *Proceedings of the Glim92 Conference and 7th International Workshop on Statistical Modelling.* New York: Springer.

Wilson P. (2007). A Hybrid Test for Non-Nested Models. In del Castillo J, Espinal A, Puig P eds. *Proceedings of the 22nd International Workshop on Statistical Modelling*, Barcelona:Universitat Autònoma de Barcelona.